Next 4 Page(s) In Document Exempt

Approved For Release 2003/09/02 : CIA-RDP80B01676R001100030012-6

TRANSLATION BY SOVIET ELECTRONIC CALCULATING MACHINE

The following is the full text of an article from PRAVDA, dated 22 January 1956 on the work of the high-speed electronic computer of the Academy of Sciences USSR. The co-authors are D. Panev, I. Mukhin and I. Bel'skaya.

The first successful experiments in the automatic translation of one language to another were conducted in the Academy of Sciences USSR. The experiments were made on the BESM high-speed electronic computing machine described in PRAVDA of 4 December 1955.

At first glance it seems unlikely that a machine can automatically translate from one language to another. But if some thought is given to the subject then it may be easily perceived that there is nothing impossible in this. In reality, language is a definite system in which the meaning of words as well as any of their variations are reflected by means of vocabulary and grammar.

Therefore, there should exist the possibility of developing dictionaries and translation rules which would account for all structural peculiarities of a sentence and which would permit the accurate, uniform determination of the essence of its component words and their relationship in the text. Thus it becomes possible to make translations by fully automatic means using machines with programming control, such as those which automatically make complex mathematical computations.

To make such a translation the sentence which is to be translated must first be coded with a special cipher code in which each letter would be represented by some 2-digit numeral. By using the Bodo code, for example where the Latin letter "a" is represented by 16, "n" by 15 and "d" by 30, we may replace the English word "and" by the numeral by 161530. The numbers corresponding to words under this codification system may be punched on a paper tape. This is done by a person who need not know the English language on a simple typewriter-type apparatus having a keyboard with Latin letters. This tape is then inserted into an electronic computing machine, in the memory device of which there is stored a previously introduced dictionary. Each word in the dictionary, consisting of an English and a Russian part, is also represented by its corresponding numeral and the process of searching for the required word in the dictionary is but a comparison of the number introduced into the machine (representing a particular word) with all the numerals of the dictionary.

E/R cy

This process may be simply represented as follows. Let the number representing the word being searched for be successively subtracted from all the numbers (each representing a word) in the dictionary.

When the result of the subtraction is equal to 0, the search is terminated: the word in the dictionary whose numerical representation coincides exactly with the numerical representation of the word being searched for has been found. This means therefore, that in the Russian portion of the dictionary there has been found the word corresponding to the English word. The machine does this completely automatically and at great speed. For instance, on the BESM machine one comparison operation takes only 1/10,000 of a second; thus the time the machine takes to look through a 1,000-word dictionary is measured in a fraction of a second.

However, the dictionary introduced into the machine differs from those dictionaries which are generally used. The dictionary must provide rules which would permit the automatic selection of the proper word from among several possibilities. To establish the meaning of a polysemantic term it is necessary to analyze its surrounding words, to study the words immediately before and after it and to study their meanings and grammatical characteristics.

Suppose we want to translate the English word "example" which has two meanings depending upon whether the word "for" precedes it or not. If the word "for" is there, the translation must be "naprimer". If the word "for" is not there, then the translation would be "primer". Considerably more complex, although similar, rules may be developed for other cases as well. These rules are formulated in the form of concrete questions to which the machine sometimes has to answer either "Yes" or "No" over 20 times. Another system of rules is needed to find English words having a particular ending in the dictionary. For instance, the word "equations" will not be found in the dictionary since it has the ending "s" indicating plurality. The dictionary lists singulars. The machine then discards the ending and the word is again run through and compared with the dictionary terms.

Finally, the dictionary must list the grammatical characteristics of its words. However, in contrast to the usual English-Russian dictionaries, it lists the grammatical characteristics not of the English but of the Russian words. This is necessary because without these characteristics it is impossible to construct a grammatically correct Russian sentence.

English grammatical characteristics are needed only insofar as they facilitate in the determination of Russian grammatical characteristics.

It is impossible to list all the grammatical characteristics in the dictionary. For example it is impossible to state in advance whether a given word would be a subject or an object. This circumstance may be determined only comparing the given word with other words and through an analysis of the entire sentence. Therefore, after using the dictionary the machine conducts a very large number of various checks of the

"Yes" and "No" answer type for the purpose of determining all the necessary grammatical characteristics of the Russian words. After this is done, a translated English phrase such as "Problems associated with Motion" reads as follows: "ZADACHA" (noun, feminine gender, second declension, plural, nominative case, soft stem ending with a sibilant), "SVYAZYVAT" (participle, past tense, feminine gender, plural, nominative), "s" (preposition, governs instrumental case), "DVIZHENIYE" (noun, neuter gender, first declension, single, instrumental case soft stem). Thus, according to the rules of Russian grammar, the translation becomes as follows: "ZADACHI, SVYAZANNYYE S DIVIZHENIYEM". The finished translation is typed on a teletype.

As an example of a translation made by this machine, the following sentence from an English text may be used: "Elementary courses in differential equations present a long list of clever devices by means of which one is supposed to be able to solve differential equations". - - "Elementarnyye kursy po differentsial'nym uravneniyam dayut dlinnyy perechen' iskusnykh priyemov, pri pomoshchi kotorykh issledovatel' kak predpolagayetsya mozhet reshat' differentsial'nyye uravneniya".

Of course it is difficult to assume that the automatic translation of literary work would be possible in the near future. Such translation is much more complex than the translation of scientific text.

The experimental work on automatic translation was conducted on the Universal computing machine "BESM", which was not specifically fitted for this purpose, by a group of scientific workers of the Academy of Sciences USSR which included, besides the authors, S. Razumovskiy, L. Korolev, N. Trifonov, G. Zelenkevich, and others. The results obtained indicate that on the basis of the principles developed there may be constructed in the future a specialized electronic machine for automatic translation.

D. Panov
I. Mukhin
I. Bel'skaya

PRAVDA, 22 January 1956

NATIONAL SCIENCE FOUNDATION
OFFICE OF THE DIRECTOR
WASHINGTON 25, D. C.

MAR 2 7 1956

Honorable Allen W. Dulles, Director
Central Intelligence Agency
Washington 25, D. C.

Dear Allen:

The development of a program of research on mechanical translation problems, as outlined in your letter of February 29, is a matter of great interest to the National Science Foundation. We agree that much is to be gained by the early development of a machine capability for translation, and we are devoting what we believe to be a reasonable fraction of our available research funds to the support of the M.I.T. linguistic studies to which you refer in your letter. We believe that an integrated attack is desirable on the problem as a whole, particularly the machine development work. Such an approach would involve careful coordination, both with respect to programming and funding.

I have asked Dr. Alberto F. Thompsen, Head of our Office of Scientific Information, to be responsible for Foundation activities in this field. He would appreciate an opportunity to meet with anyone you care to designate to consider the problem of appropriate levels and sources of support in the Government for work in this field. We assume that other agencies should be consulted; for example, we understand that the Air Force is undertaking a major effort in machine development for translation.

The Foundation is prepared to administer any part of a program of research in machine translation which is agreed by all concerned to be desirable. As you know, it is our policy to rely heavily on the guidance of advisory panels of experts in each field of research to advise us on the handling of research proposals directed to the Foundation. We are in the process of organizing such a group at the present time which we feel would be fully capable of guiding the progress of machine translation research.

I hope that the kind of approach to the problem outlined in this letter appears reasonable, and I look forward to the development of an effective program in this field.

Sincerely yours,

Alan

Alan T. Waterman
Director

ER-7-9554/a

MAR 3 1956

Mr. Alan T. Waterman
Director
National Science Foundation
Washington 25, D. C.

Dear Alan:

Thank you very much for the copies of the Fifth Annual Report of the National Science Foundation and the National Science Foundation Act forwarded with your letter of 23 February.

I have reviewed these papers with much interest and have referred the additional copies to others here for study. They should be of considerable value in our future activities.

Sincerely,

Allen W. Dulles
Director

STAT O/DCI/[redacted]hc (28 Feb 56)
Distribution:
Orig - Addressee
1 - DCI
STAT 1 - [redacted](Chrono)
1 - ER w/basic
1 - Reading

NATIONAL SCIENCE FOUNDATION
OFFICE OF THE DIRECTOR
WASHINGTON 25, D. C.

February 23, 1956

PERSONAL

The Honorable Allen W. Dulles
Director
Central Intelligence Agency
Washington 25, D. C.

Dear Allen:

At our recent meeting I promised to send you a copy of the Fifth Annual Report of the National Science Foundation. Herewith are three in case other members of your staff may find the report useful.

I also enclose a copy of the National Science Foundation Act which gives the formal basis for our interest and authority in the field of international science activities.

I believe the subject we discussed is an important and timely one, and I am looking forward with great pleasure to working with you on ways and means to make progress.

With best regards,

Sincerely yours,

Alan

Alan T. Waterman
Director

Enclosures

1 April 1957

Dr. Alan T. Waterman
Director
National Science Foundation
1520 H Street, N. W.
Washington, D. C.

Dear Alan:

Many thanks for sending me a copy of "Science in the Federal Government." I am glad to be able to add this to my library on this general subject.

One of these days we should have a talk together about the scientific attache question. I have been pressing STAT [redacted] on the subject and I know he is doing what he can but progress is slow.

Sincerely,

Allen W. Dulles
Director

AWD/ji
Distribution:
1 - DCI
1 - ER
1 - Reading

Next 1 Page(s) In Document Exempt

Approved For Release 2003/09/02 : CIA-RDP80B01676R001100030012-6

TRANSLATION BY SOVIET ELECTRONIC CALCULATING MACHINE

The following is the full text of an article from PRAVDA, dated 22 January 1956 on the work of the high-speed electronic computer of the Academy of Sciences USSR. The co-authors are D. Panev, I. Mukhin and I. Bel'skaya.

The first successful experiments in the automatic translation of one language to another were conducted in the Academy of Sciences USSR. The experiments were made on the BESM high-speed electronic computing machine described in PRAVDA of 4 December 1955.

At first glance it seems unlikely that a machine can automatically translate from one language to another. But if some thought is given to the subject then it may be easily perceived that there is nothing impossible in this. In reality, language is a definite system in which the meaning of words as well as any of their variations are reflected by means of vocabulary and grammar.

Therefore, there should exist the possibility of developing dictionaries and translation rules which would account for all structural peculiarities of a sentence and which would permit the accurate, uniform determination of the essence of its component words and their relationship in the text. Thus it becomes possible to make translations by fully automatic means using machines with programming control, such as those which automatically make complex mathematical computations.

To make such a translation the sentence which is to be translated must first be coded with a special cipher code in which each letter would be represented by some 2-digit numeral. By using the Bodo code, for example where the Latin letter "a" is represented by 16, "n" by 15 and "d" by 30, we may replace the English word "and" by the numeral by 161530. The numbers corresponding to words under this codification system may be punched on a paper tape. This is done by a person who need not know the English language on a simple typewriter-type apparatus having a keyboard with Latin letters. This tape is then inserted into an electronic computing machine, in the memory device of which there is stored a previously introduced dictionary. Each word in the dictionary, consisting of an English and a Russian part, is also represented by its corresponding numeral and the process of searching for the required word in the dictionary is but a comparison of the number introduced into the machine (representing a particular word) with all the numerals of the dictionary.

This process may be simply represented as follows. Let the number representing the word being searched for be successively subtracted from all the numbers (each representing a word) in the dictionary.

When the result of the subtraction is equal to 0, the search is terminated: the word in the dictionary whose numerical representation coincides exactly with the numerical representation of the word being searched for has been found. This means therefore, that in the Russian portion of the dictionary there has been found the word corresponding to the English word. The machine does this completely automatically and at great speed. For instance, on the BESM machine one comparison operation takes only 1/10,000 of a second; thus the time the machine takes to look through a 1,000-word dictionary is measured in a fraction of a second.

However, the dictionary introduced into the machine differs from those dictionaries which are generally used. The dictionary must provide rules which would permit the automatic selection of the proper word from among several possibilities. To establish the meaning of a polysemantic term it is necessary to analyze its surrounding words, to study the words immediately before and after it and to study their meanings and grammatical characteristics.

Suppose we want to translate the English word "example" which has two meanings depending upon whether the word "for" precedes it or not. If the word "for" is there, the translation must be "naprimer". If the word "for" is not there, then the translation would be "primer". Considerably more complex, although similar, rules may be developed for other cases as well. These rules are formulated in the form of concrete questions to which the machine sometimes has to answer either "Yes" or "No" over 20 times. Another system of rules is needed to find English words having a particular ending in the dictionary. For instance, the word "equations" will not be found in the dictionary since it has the ending "s" indicating plurality. The dictionary lists singulars. The machine then discards the ending and the word is again run through and compared with the dictionary terms.

Finally, the dictionary must list the grammatical characteristics of its words. However, in contrast to the usual English-Russian dictionaries, it lists the grammatical characteristics not of the English but of the Russian words. This is necessary because without these characteristics it is impossible to construct a grammatically correct Russian sentence.

English grammatical characteristics are needed only insofar as they facilitate in the determination of Russian grammatical characteristics.

It is impossible to list all the grammatical characteristics in the dictionary. For example it is impossible to state in advance whether a given word would be a subject or an object. This circumstance may be determined only comparing the given word with other words and through an analysis of the entire sentence. Therefore, after using the dictionary the machine conducts a very large number of various checks of the

"Yes" and "No" answer type for the purpose of determining all the necessary grammatical characteristics of the Russian words. After this is done, a translated English phrase such as "Problems associated with Motion" reads as follows: "ZADACHA" (noun, feminine gender, second declension, plural, nominative case, soft stem ending with a sibilant), "SVYAZYVAT" (participle, past tense, feminine gender, plural, nominative), "s" (preposition, governs instrumental case), "DVIZHENIYE" (noun, neuter gender, first declension, single, instrumental case soft stem). Thus, according to the rules of Russian grammar, the translation becomes as follows: "ZADACHI, SVYAZANNYYE S DIVIZHENIYEM". The finished translation is typed on a teletype.

As an example of a translation made by this machine, the following sentence from an English text may be used: "Elementary courses in differential equations present a long list of clever devices by means of which one is supposed to be able to solve differential equations". - - "Elementarnyye kursy po differentsial'nym uravneniyam dayut dlinnyy perechen' iskusnykh priyemov, pri pomoshchi kotorykh issledovatel' kak predpolagayetsya mozhet reshat' differentsial'nyye uravneniya".

Of course it is difficult to assume that the automatic translation of literary work would be possible in the near future. Such translation is much more complex than the translation of scientific text.

The experimental work on automatic translation was conducted on the Universal computing machine "BESM", which was not specifically fitted for this purpose, by a group of scientific workers of the Academy of Sciences USSR which included, besides the authors, S. Razumovskiy, L. Korolev, N. Trifonov, G. Zelenkevich, and others. The results obtained indicate that on the basis of the principles developed there may be constructed in the future a specialized electronic machine for automatic translation.

D. Panov
I. Mukhin
I. Bel'skaya

PRAVDA, 22 January 1956

STAT

SENDER WILL CHECK CLASSIFICATION TOP AND BOTTOM

UNCLASSIFIED | CONFIDENTIAL | SECRET

CENTRAL INTELLIGENCE AGENCY

OFFICIAL ROUTING SLIP

TO	NAME AND ADDRESS	INITIALS	DATE
1			
2			
3			
4			
5			
6			

	ACTION		DIRECT REPLY		PREPARE REPLY
	APPROVAL		DISPATCH		RECOMMENDATION
	COMMENT		FILE		RETURN
	CONCURRENCE		INFORMATION		SIGNATURE

Remarks:

Pls prepare reply for the Director's signature.

(The Director has one copy of the report. He evidently kept copy of NSF Act mentioned in para 2.)

FOLD HERE TO RETURN TO SENDER

FROM: NAME, ADDRESS AND PHONE NO.	DATE
per JSE 24 Feb	

UNCLASSIFIED | CONFIDENTIAL | SECRET

FORM NO. 237 Replaces Form 30-4 which may be used. (40)
1 APR 55
U. S. GOVERNMENT PRINTING OFFICE : 1955—O-342531